International Journal for Modern Trends in Science and Technology, 8(07): 170-175, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0807025
Available online at: http://www.ijmtst.com/vol8issue07.html

Social Distancing and Face Mask Detection in Real-Time Using Deep Learning

Kirti Mishra | Kavita Kelkar

Computer Engineering, KJSCE, Mumbai, Maharashtra, India
Email: kirti.dm@somaiya.edu

To Cite this Article
Kirti Mishra and Kavita Kelkar. Social Distancing and Face Mask Detection in Real-Time Using Deep Learning. International Journal for Modern Trends in Science and Technology 2022, 8(07), pp. 170-175. https://doi.org/10.46501/IJMTST0807025

Article Info
Received: 05 June 2022; Accepted: 01 July 2022; Published: 09 July 2022.

ABSTRACT

Humanity is suffering greatly as a result of the coronavirus sickness, which appeared in 2020. Wearing a "Face Mask" in public and upholding "Social Distancing" are the only preventative measures we have against this pandemic. Additionally, a lot of service providers, including hotels, hospitals, train stations, and airports, require customers to use the service only if they appropriately wear the Mask and maintain social distance. Because it would take a lot of labor, it would be impossible to personally check people to see if they adhered to the social distance and mask requirement.

In this paper, people are detected using pretrained Yolov3, and social distance violations are determined using Euclidean distance. Deep learning modules will be applied for face mask detection. Without a manual monitoring system, the system offers safety and security tools. It can be used in any hospital, office, public building, educational institution, work site, airport, etc. The proposed face mask and social distance detection system has the potential to be used to help assure individual and group safety if implemented properly.

KEYWORDS:YOLO, Image Processing, Object Detection, CNN, Deep Learning, Bounding Boxes, Neural Network

1. INTRODUCTION

Colds to fatal diseases like the Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) can all be brought on by the large group of viruses known as coronaviruses [1]. 90,054,813 confirmed cases of COVID-19 worldwide, including 1,945,610 fatalities, have been recorded to WHO as of January 2021 [2]. Direct contact with infected people can spread the COVID-19 virus, as can indirect contact (contaminated environment). Respiratory droplets, which are droplet particles with diameters between 5 and 10 m, cause this. The risk of having one's mouth, nose, or eyes exposed to potentially infectious respiratory droplets exists when a person is in close proximity (within 1 m) to someone who is experiencing respiratory symptoms [3]. It is essential to limit the virus's spread as much as possible given the rise in infected cases and fatalities.

To stop the spread of COVID-19, social distance should be used in conjunction with other routine preventive measures such mask use, refraining from touching your face with unwashed hands, and periodically washing your hands with soap and water for at least 20 seconds. The act of maintaining a secure

space between oneself and those around oneself is known as social distancing [4]. The recommended distance to prevent the virus from spreading by touch is two meters (six feet) [5,6]. Following the social distance guidelines reduced interaction by 95% for people over the age of 60 and by 85% for people under the age of 20 [7]. This demonstrates the potential for "flattening the curve" by adhering to the proper social distance rules.

Respiratory droplets are the main method of transmission for COVID-19. Coughing, sneezing, talking, shouting, or singing all cause respiratory droplets to enter the air. People may then breathe in these droplets or have them land in their mouths or nostrils. Thus, wearing a mask is necessary to stop the infection from spreading. Simple barriers like masks can help stop respiratory droplets from getting to other people. According to studies, when worn over the mouth and nose, masks lessen the number of droplets that spray. [8,9]

With limited resources, manually observing social distance rules and looking at people's face masks can be both time-consuming and error-prone. The public must immediately understand the ideal social distance norms that should be practiced in order to control the virus' spread. This includes detecting social distance violations and categorizing face masks in order to assess the safety of the populace by determining whether adequate distance is maintained and whether face masks are worn. This system can be used in a variety of public locations using cameras, including supermarkets, gas stations, and traffic lights. This offers processing methods that can be used with a surveillance system for many additional purposes.

In order to prevent the spread of the COVID-19 virus, this study suggests a proactive and portable surveillance system that can monitor people's movements and alert authorities to any suspicious activity.

2. Literature Review

Deep transfer learning (ResNet-50) and conventional machine learning techniques are combined in the model developed by Mohamed Loey et al. [10]. To enhance the performance of the ResNet-50 model, the final layer was eliminated and replaced with three conventional machine learning classifiers (Support vector machine (SVM), decision tree, and ensemble). One dataset, out of the four different kinds they used, comprised the most photos overall and took the longest to train compared to the others since it included both real and fake face masks. Additionally, there is no reported accuracy for this sort of dataset in accordance with related publications. The decision trees classifier failed to achieve a good classification accuracy (68 percent) on the training dataset with genuine face masks.

A backbone, neck, and heads detection network with Resnet as the backbone, FPN (feature pyramid network) as the neck, and classifiers, predictors, estimators, etc. as the heads is implemented [11]. However, it is challenging for learning algorithms to learn improved features because of the small size of the face mask dataset. Better detection accuracy must be attained because there is less study on face mask detection.

Another method is RinkalKeniya's self-created SocialdistancingNet-19 model, which displays labels to determine whether a person is safe or unsafe if the distance is smaller than a predetermined threshold. People must be moving continually if a webcam is to be used; otherwise, the detection will be inaccurate. This may occur as a result of the network's detection mechanism, which involves detecting the full frame and using centroids to calculate the distance between individuals (brute force approach).

For social distance and mask detection, Shashi Yadav developed a deep learning strategy utilizing Single Shot object Detection (SSD) with MobileNet V2 and OpenCV [13]. The difficulty with this method is that it labels persons as masked if they have their hands over their faces or their faces are obscured by things. These examples are inappropriate for this model. Although an SSD may detect many things in a frame in this case, it can only detect one human in this system.

The majority of the publications have addressed either the problem of face mask detection or social distancing monitoring. Additionally, there is still room to use stronger models to gain greater accuracy where both were used. In our study, we discuss the significance of prediction time, a feature that is absent in other papers, and we use prediction time as an assessment metric, which is essential for the system's practical application. The model suggested by the paper for person detection is YOLOv3, a cutting-edge object detection model, followed by DBSCAN to calculate the distances between people and perform clustering to

identify whether or not they are far apart. This approach is significantly superior to other clustering methods like brute force distance calculations or k-means, which require that the number of clusters be determined before performing clustering.

We have DSFD, a potent feature extractor with decent face detection accuracy, for face detection. Additionally, MobileNetV2 was employed for face mask categorization since it outperformed Xception and ResNet50 in terms of performance. A labelled video dataset was also developed for testing the system by labelling the video frames, and a mask dataset was created utilizing data augmentation techniques.

3. METHODOLOGY

By automatically observing individuals to see if they keep a safe social distance and by identifying persons who are wearing face masks, the suggested system contributes to ensuring the safety of people in public spaces. This section briefly explains the proposed system's architecture and how it will work automatically to stop the coronavirus from spreading.

To automatically monitor people in public spaces with a camera built into a local device and to identify persons wearing masks or not, the proposed system combines a transfer learning method to performance optimization with a deep learning algorithm and computer vision. In addition to feature extraction, we also perform fine tuning, which is an additional transfer learning method. In this process, camera video feeds from the Network Video Recorder (NVR) are streamed using RTSP. To increase speed and accuracy, these frames are then transformed to grayscale and sent to the model for additional processing inside the machine. We chose the RestNet50 architecture as the foundational model for detection because it offers a significant cost advantage over the traditional 2D CNN model. The YOLOV3Detector, a neural network architecture that has already been honed for high-quality picture classification on a big collection of photos like ImageNet and Pascal, is also a part of the process.

The RestNet50 is being loaded with pre-trained ImageNet weights, the network head is being left off, a new FC head is being built, attached to the base in place of the old head, and the network's base layers are being frozen. During the backpropagation's fine-tuning stage, the head layer weights will be modified but not the weights of these base layers. The model is built and trained after the data is ready and the model architecture is set up for fine tuning. Experiments have been conducted with OpenCV, Keras using Deep Learning and Computer Vision to examine the safe social distance between detected persons and face Mask detection in real-time video streams. A very small learning rate is used during the retraining of the architecture to ensure that the convolutional filters already learned do not deviate dramatically. The three components of the proposed system—person detection, measuring the safe distance between identified persons, and face mask detection—are its key contributions. Real-time person detection is accomplished with the aid of YOLOV3 (You Only Look Once), which outperforms the comparable state-of-the-art Faster R-CNN model by achieving 91.2 percent mAP utilizing RestNet50 and OpenCV. Every person that is found will have a bounding box displayed around them. In this system, YOLOV3 is only able to detect one person even though it can recognize numerous things in a frame. To determine the distance between two people, we must first determine the person's distance from the camera using the triangle similarity technique, determine the camera's perceived focal length, assume that the person's distance from the camera is D, and that the person's actual height is H, which is 165 cm, and then use the YOLOV3 person detection algorithm to determine the person's pixel height P using the bounding box coordinates. The following formula can be used to determine the camera's focal length using these values:

$$F = (P \times D) / H$$

Then, we calculate the real person's distance from the camera using the real person's height H, the real person's pixel height P, and the real person's focal length F. The following methods can be used to calculate your distance from the camera:

$$D1 = (H \times F) / P$$

We determine the distance between two persons in the video after determining the depth of the person in the camera. In a video, several persons may be seen. The midpoint of the bounding boxes of all detected persons are thus measured using the Euclidean distance. As a result, we obtained the x and y values, which we then converted into centimeters. We know each person's x, y, and z coordinates in centimeters, which represent their

distance from the camera. Using the (x, y, z) coordinates, the Euclidean distance between each person found is determined. A red bounding box is displayed around two people if their separation is smaller than 2 meters or 200 cm, signifying that they do not maintain a social distance. In the suggested system, transfer learning is applied on top of the highly effective, previously trained YOLOV3 model for face identification, which is supported by the RestNet50 architecture, to provide a compact, accurate, and computationally efficient model that is simpler to deploy to a machine. We used unique face crop datasets of 3165 or so photos, both with and without mask annotations. Annotated images are used to train a deep learning binary classification model, which uses output class confidence to classify the input image into the mask and no mask categories. The output of the YOLOV3model displays a bounding box and extracts a human mask. The proposed system continuously monitors public spaces, and when a person without a mask is seen, the person's face is captured and an alert sent to the authorities with a face image. At the same time, the distance between people is measured in real time, and if more than 20 people have been continuously found to be violating safe social distance standards at the threshold time, an alert is sent to the control center at the State Police Headquarters to take further action. Due to the Covid-19 outbreak, this system can be employed in real-time applications that call for the secure monitoring of social distances between individuals as well as the detection of face masks for security reasons. We chose to utilize this architecture because deploying our model to edge devices for continuous monitoring of public spaces could ease the strain of physical monitoring. To ensure that public safety regulations are obeyed, this system can be connected with edge devices for usage in airports, train stations, offices, schools, and other public spaces.

A. Person Detection

For person detection, the YOLOv3 model was utilized [16]. The 106-layered fully convolutional neural network is made up of 53 layers of Darknet-53 trained on ImageNet, which serves as a potent feature extractor, and an additional 53 layers for detection. The YOLOv3 architecture is shown in Fig. 5. Utilized is an anchor box with three scales: 13x13, 26x26, and 52x52. As indicated in picture, the presence of a person is predicted using these three anchor boxes. Following prediction, this model outputs a list of bounding boxes together with the degree of confidence in the detected person class. The problem of overlapping bounding boxes resulting in numerous detections of the same object is resolved by non-maximum suppression (NMS). Based on the confidence value and NMS threshold, which had values of 0.5 and 0.3 respectively, the final bounding boxes were chosen. As a result, only classes with a confidence level of at least 50% are included, and bounding boxes with an overlap of more than 30% are all disregarded.

B. Social Distancing

The DBSCAN algorithm was used to determine whether the social distance between the individuals was still present. Similar points are grouped together via an unsupervised learning technique. The number of clusters need not be predetermined before training with DBSCAN, in contrast to the k-means algorithm. While forming the clusters, it also ignores the noisy or outlier data.

Since social distance is assessed between a minimum of two individuals, the cluster's minimum required points were set at two, and the distance parameter was 200. If, when taking each individual into account, the distance between them is smaller than the distance parameter, they are clustered together into a cluster. A person is labelled as safe and bound with orange boxes if they do not belong to any clusters. Red lines between individuals in a cluster are used to indicate that they are too near to one another, and blue boxes are used to connect those individuals.

C. Face Detection

In terms of accuracy and prediction time, the two pretrained models for face identification are DSFD and RetinaNetMobileNetV1. RetinaNetMobileNetV1 is a compact single-shot face detector that was initially created for the face detection models' mobile deployment.

The first shot detector, which is made up of convolutional layers, the feature enhance module, which creates additional features, and the second shot detector, which combines the enhanced features with first shot detector loss to produce the final predictions, make up the DSFD architecture. The model works significantly better than a single shot detector because a second shot is used, although prediction is extremely slow.

DSFD outperforms RetinaNetMobileNetv1 in terms of accuracy but with a substantially longer detection time. The faces are inevitably hazy and blurry because the photographs from the video frames were taken from the security camera. Because accuracy cannot be compromised, the DSFD model was selected.

D. Face Mask Classification

To determine whether a mask is present on the faces discovered, face mask classification is accomplished using CNN binary image classification architecture. The performance of several models for classifying masks on 128x128 images was compared for class 0 (no mask) and class 1 (masked) in terms of accuracy, precision, recall, and F1 scores as shown in Table I. Due to its accuracy and forecast time performance, MobileNetV2 was selected.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Resnet50	47.7	49, 46	48, 57	49, 47
Xception	50.8	51, 51	46, 56	48, 53
MobilenetV2	93.2	94.6, 93.80	95.7, 94.1	95.1, 93.9

Table I: Comparison between face mask classifier models

4. RESULT

Below equations were used to determine the accuracy and F1 score. Here, TP stands for the total number of true positives, TN for true negatives, FP for false positives, and FN for false negatives.

$$\text{Accuracy} = \frac{TP+TN}{(TP+FP)+(TN+FN)}$$

$$\text{F1 Score} = 2 * \frac{Precision*Recall}{(Precision + Recall)}$$

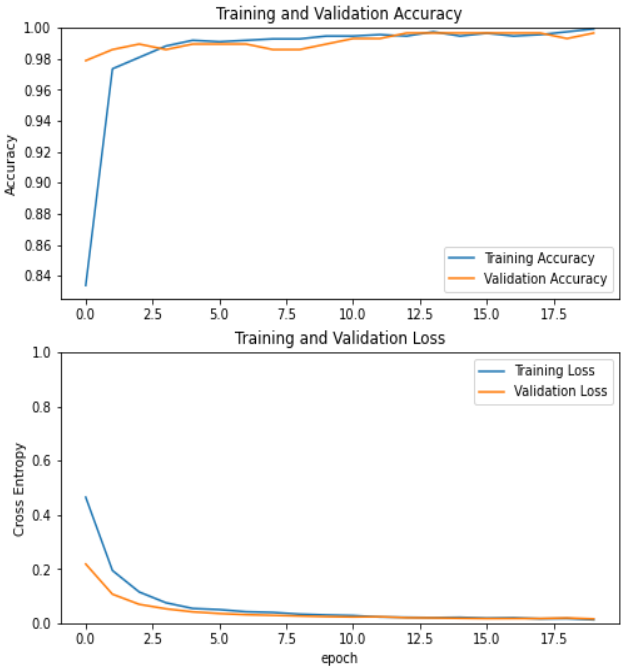

Fig 1 (A): Training & Validation Accuracy
(B): Training & Validation Loss

5. CONCLUSION

In public spaces where it is exceedingly challenging to manually monitor social distance behaviors, this study offers an effective approach. The YoloV3 algorithm can be used in a variety of sectors to address certain real-world issues, including security, lane monitoring, and even providing aural feedback to help the blind. In this, a model is developed to recognize persons instantly. Additionally, the database used to train the model might be bigger and more varied in order to raise the mAP while lowering the model's final average losses. As a result, the model is able to identify objects even in murky, complicated settings. Training can be completed much more quickly, therefore there is space for improvement by including a more potent and quicker GPU. This method can be used in a variety of settings to address practical issues like security, lane monitoring, or even providing aural feedback to help the blind.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] US Centers for Disease Control and Prevention. "Interim pre-pandemic planning guidance: community strategy for pandemic influenza mitigation in the United States: early, targeted, layered use of nonpharmaceutical interventions Atlanta: The Centers; 2007." https://stacks.cdc.gov/view/cdc/11425 [Online; accessed 13 Jan 2021]

[2] World Health Organization (WHO) "WHO Coronavirus Disease (COVID-19) Dashboard" https://covid19.who.int/, [Online; accessed 13 Jan 2021].

[3] World Health Organization (WHO) "Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations "https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations [Online; accessed 13 Jan 2021]

[4] Center for disease control and prevention (CDC) "COVID-19 Social distancing"https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting- sick/social-distancing.html [Online; accessed 13 Jan 2021]

[5] Manasee Mishra, Piyusha Majumdar; Social Distancing During COVID-19: Will it Change the Indian Society? (2020)

[6] Marco Cristan, Alessio Del Bue, Vittorio Murino, FraccescoSetti And Alessandro Vinciarelli The Visual Social Distancing Problem, 2020

[7] Matrajt L, Leung T. Evaluating the efficacy of social distancing strategies to postpone or flatten the curve of coronavirus disease. Emerg Infect Dis, man (2020)

[8] [8] World Health Organisation (WHO) "Coronavirus disease (COVID- 19) advice for the public: When and how to use masks" https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for- public/when-and-how-to-use-masks [Online; accessed 13 Jan 2021]

[9] Center for disease control and prevention (CDC) "Considerations for Wearing Masks" https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html [Online; accessed 13 Jan 2021]

[10] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa; A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic (2020)

[11] Mingjie Jiang*, Xinqi Fan*, Hong Yan, RETINAFACEMASK: A FACE MASK DETECTOR (2020)

[12] Shashi Yadav, Goel Institute of Technology and Management, Dr. A.P.J. Abdul Kalam Technical University, Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence (2020)

[13] RinkalKeniya · NinadMehendale, Real-time social distancing detector using Socialdistancing-Net19 deep learning network (2020)

[14] Indhu Jain, Mr. Sudhir Goswami; A Comparative Study of Various Image Restoration techniques with different types of blur, International Journal Of Research In Computer Applications And Robotics (2015).

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; Deep Residual Learning for Image Recognition (2015).

[16] Joseph Redmon, Ali Farhadi from University of Washington; YOLOv3: An Incremental Improvement. (2018)

[17] "MobileNetV2: Inverted Residuals and Linear Bottlenecks", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018